

Few paths, fewer words: model selection with automatic structure functions

Bjørn Kjos-Hanssen

August 5, 2016

Abstract

We consider the problem of finding an optimal statistical model for a given binary string. Following Kolmogorov, we use structure functions. In order to get concrete results, we replace Turing machines by finite automata and Kolmogorov complexity by Shallit and Wang's automatic complexity.

The p -value of a model for given data x is the probability that there exists a model with as few states, accepting as few words, fitting uniformly randomly selected data y .

Deterministic and nondeterministic automata can give different optimal models. For $x = 01111011011$, the best deterministic model has p -value 0.3, whereas the best nondeterministic model has p -value 0.04.

In the nondeterministic case, counting paths and counting words can give different optimal models. For $x = 0110001000$, the best path-counting model has p -value 0.79, whereas the best word-counting model has p -value 0.60.

1 Introduction

Shallit and Wang [6] introduced *automatic complexity* (defined below) as a computable alternative to Kolmogorov complexity. They considered deterministic automata, whereas Hyde and Kjos-Hanssen [2] studied the nondeterministic case, which in some ways behaves better.

Unfortunately, even nondeterministic automatic complexity is somewhat inadequate. The word 00010000 has maximal nondeterministic complexity among all binary strings of length 8. However, intuitively it is quite simple. One way to remedy this situation is to consider a *structure function* analogous to that for Kolmogorov complexity. The latter was introduced by Kolmogorov at a 1973 meeting in Tallinn and studied by Vereshchagin and Vitányi [8], Rissanen [5], and Staiger [7].

Here we show that some notions in this area, in the non-deterministic setting, depend on whether we are counting accepting words or accepting paths. This is interesting because counting words is most efficient for compression, whereas counting paths seems to lead to more time-efficient computability.

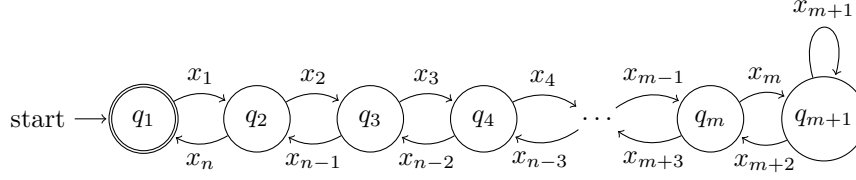


Figure 1: A nondeterministic finite automaton that only accepts one word $x = x_1x_2x_3x_4 \dots x_n$ of length $n = 2m + 1$.

Several results are proved by computer. We do not know of short computer proofs (certificates) in most cases, so we only include the claim that the result was proved by computer.

We let $L(M)$ denote the language recognized by the automaton M .

Definition 1 (Shallit and Wang [6]). *The automatic complexity of a finite binary word $x = x_1 \dots x_n$ is the least number $A(x)$ of states of a deterministic finite automaton M such that*

$$L(M) \cap \{0, 1\}^n = \{x\},$$

that is, x is the only word of length n accepted by M . If we do not require the transition function of M to be total, we obtain the nontotal automatic complexity $A^-(x)$.

We will consider model selection in three distinct *modes*:

1. the deterministic mode δ ,
2. the path-counting nondeterministic mode π , and
3. the word-counting nondeterministic mode ω .

Formally, we could take $\{1, 2, 3\} = \{\delta, \pi, \omega\}$.

Definition 2. *The number of acceptances $\text{Acc}_n^\mu(M)$ at length n for an NFA M in mode μ is defined as follows.*

- *If μ is the deterministic mode then $\text{Acc}_n^\mu(M)$ is ∞ (or undefined) if M is not deterministic. If M is deterministic then $\text{Acc}_n^\mu(M)$ is the number of words of length n accepted by M .*
- *If μ is the path-counting nondeterministic mode then $\text{Acc}_n^\mu(M)$ is the number of paths of length n leading to an accept state of M .*
- *If μ is the word-counting nondeterministic mode then $\text{Acc}_n^\mu(M)$ is the number of words of length n accepted by M .*

Following Kolmogorov, we shall rarely consider more fine-grained acceptance counting than just by powers of b . So we define $\log \text{Acc}_n^\mu(M, b) = \lceil \log_b \text{Acc}_n^\mu(M) \rceil$.

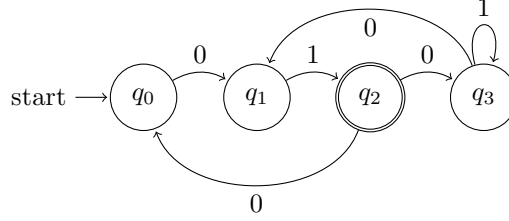


Figure 2: An automaton accepting only $x = 010111010$ along exactly one path, but accepting other words along multiple paths, giving Theorem 7.

Definition 3 ([2]). *The path-counting nondeterministic automatic complexity $A_N^\pi(w)$ of a word w is the minimum number of states of an NFA M such that M accepts w and $\text{Acc}_{|w|}^\pi(M) = 1$.*

We assume that our NFAs are not generalized, i.e., they have no ε -transitions.

Theorem 4. *It does not matter for A_N^π or A_N^ω whether ε -transitions are allowed.*

Proof. Given an automaton M using ε -transitions, we define another automaton M' not using any ε -transitions. We put a transition in M' between states q_1 and q_2 labeled i if there is some path from q_1 to q_2 in M whose labels concatenate to i under the obvious rule that $i\varepsilon = i = \varepsilon i$. \square

We assume our automata have only a single accept state.

The definition of A_N^π is not robust under permutation of quantifiers, in the following sense.

Definition 5. *Let $A^\dagger(x)$ be the minimum number of states of an NFA such that x is the only string of length n that is accepted along exactly one path (but other strings may be accepted among more than one path).*

When considering automatic complexity it is often sufficient to replace an automaton by a *state sequence*. A state sequence is a sequence of states, typically the sequence of states visited by the automaton during the processing a word. For computational purposes we may represent a state sequence q_0, \dots, q_n as a sequence of nonnegative integers $s = s_0 \dots s_n$ with the property that $s_i \leq \max_{j < i} s_j + 1$.

Theorem 6 (proved by computer). $A_N^\pi(010111010) = 5$.

Theorem 7. *There is an x such that $A^\dagger(x) < A_N^\pi(x)$.*

Proof. Consider $x = 010111010$. By Theorem 6, $A_N^\pi(x) = 5$. As the state sequence 0123333120 witnesses, $A^\dagger(x) \leq 4$. (See Figure 2.) \square

Definition 8. *Let $n = 2m + 1$ be a positive odd number, $m \geq 0$. A finite automaton of the form given in Figure 1 for some choice of symbols x_1, \dots, x_n and states q_1, \dots, q_{m+1} is called a Kayleigh graph.*

Theorem 9 (Hyde [2]). *The nondeterministic automatic complexity $A_N^\pi(x)$ of a word x of length n satisfies*

$$A_N^\pi(x) \leq \lfloor n/2 \rfloor + 1.$$

Proof. If n is odd, then a Kayleigh graph witnesses this inequality. If n is even, a slight modification suffices. \square

Definition 10. *Let $\mu \in \{\omega, \pi\}$. Suppose x is a binary word of length n . S_x^μ is defined to be the set of pairs of integers (q, m) such that there exists an NFA M with $x \in L(M)$, at most q states, and $\log \text{Acc}_n^\mu(M, 2) \leq m$.*

We note that S_x^μ has the upward closure property

$$q \leq q', m \leq m', (q, m) \in S_x^\mu \implies (q', m') \in S_x^\mu.$$

From S_x^μ we can define the structure function h_x^μ and the dual structure function $h_x^{*\mu}$. The definition was presented to us by Vereshchagin (personal communication, 2014), inspired by [8].

Definition 11.

$$\begin{aligned} h_x^{*\mu}(m) &= \min\{k : (k, m) \in S_x^\mu\}, \\ h_x^\mu(k) &= \min\{m : (k, m) \in S_x^\mu\}. \end{aligned}$$

Note that $S_x^\pi \subseteq S_x^\omega$ and hence $h_x^\omega(k) \leq h_x^\pi(k)$ and $h_x^{*\omega}(m) \leq h_x^{*\pi}(m)$ for each k, m , and x . Upper bounds on $h_x^{*\omega}(m)$, generalizing the $m = 0$ case covered by Hyde's Theorem 9, were studied in [4]. It may be observed that the proofs given there are based on counting accepting paths and hence apply equally to $h_x^{*\pi}$.

Definition 12. *For a word x , the word-based nondeterministic automatic complexity of x is defined by*

$$A_N^\omega(x) = h_x^{*\omega}(0).$$

Equivalently,

$$A_N^\omega(x) = A_N^\omega(\{x\}).$$

Conjecture 13. *There is an x such that $A_N^\omega(x) \neq A_N^\pi(x)$.*

Conjecture 13 lies at the crossroads of, and indeed was the inspiration for, the following results.

- Theorem 19: $h_x^{*\omega}(m) \neq h_x^{*\pi}(m)$ for the word 000010000, at $m = 1$. Conjecture 13 states that there is even such an example with $m = 0$.
- Theorem 30: $A_N^\omega(\mathcal{F}) \neq A_N^\pi(\mathcal{F})$ for the doubleton $\mathcal{F} = \{0110, 1111\}$. Conjecture 13 states that there is a singleton example.

Theorem 14 (proved by computer). *There is no binary word x with $|x| \leq 10$ and $A_N^\omega(x) \neq A_N^\pi(x)$.*

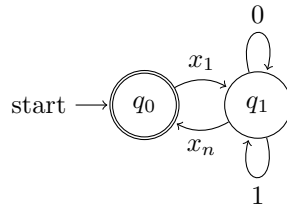
2 Structure functions

We now show that the automatic complexity structure function of a word x sometimes depends on whether we are counting accepting paths or accepted words.

Theorem 15. *For any word $x = x_1 \dots x_n$,*

$$h_x^{*\omega}(n-2) \leq 2.$$

Proof. It suffices to consider the following NFA:



□

Theorem 16 (proved by computer). *Let $x = 001011$. Then $h_x^{*\pi}(|x| - 2) = 3$.*

Theorem 17. *There is a binary word x of length 6 and an m such that $h_x^{*\omega}(m) < h_x^{*\pi}(m)$.*

Proof. Let $x = 001011$. By Theorem 15 and Theorem 16,

$$h_x^{*\omega}(|x| - 2) \leq 2 < 3 = h_x^{*\pi}(|x| - 2). \quad \square$$

Theorem 17 is optimal for n , but not for m , as Theorem 19 indicates.

Theorem 18 (proved by computer). *Let $x = 000010000$. Then $h_x^{*\pi}(1) = 5$.*

Theorem 19. *There is a word x such that*

$$h_x^{*\omega}(1) < h_x^{*\pi}(1).$$

Proof. Let $x = 000010000$. By Figure 3, $h_x^{*\pi}(1) \leq 4$. Hence by Theorem 18,

$$h_x^{*\pi}(1) \leq 4 < 5 = h_x^{*\omega}(1). \quad \square$$

3 Model selection

The automatic structure functions are intended to provide statistical explanations for words. The best explanation for a word x is the automaton witnessing a value of the structure function that is unusually low, compared to other words y . It turns out that the phenomenon of Theorem 17 also applies to such best explanations.

As envisioned by Kolmogorov, structure functions have potential applications in computational statistics. We now describe concrete results of our foray into model selection with structure functions for automatic complexity.

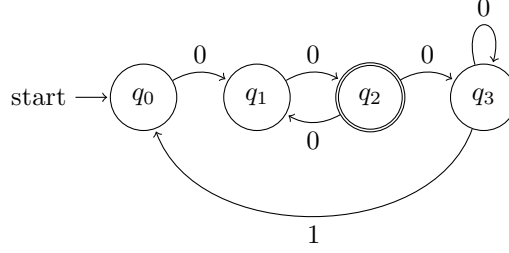


Figure 3: An automaton accepting $x = 0^410^4$ along one path and 0^610^2 along two paths, giving Theorem 19.

Definition 20. Let X be a uniform random variable on $[b]^n$. The b -ary p -value achieved by an NFA M at a length n in mode μ is the probability that X is accepted by some NFA N such that N has no more states than M , and $\log \text{Acc}_n^\mu(N, b) \leq \log \text{Acc}_n^\mu(M, b)$.

Definition 21. An NFA M is an optimal b -ary model for x in mode μ if M accepts x and M achieves the minimal b -ary p -value at length $|x|$ in mode μ among all NFAs that accept x . The b -ary explanation of x in mode μ is the set of all optimal b -ary models for x in mode μ .

Often, we take b to be the least integer such that x is a word in the alphabet $[b]$. For binary words, we usually take $b = 2$, even in the case of the word 0^n .

Theorem 22 (proved by computer). Let $x = 01111011011$. In both the path-counting mode and the deterministic mode the optimal number of states for x is 3. The only optimal state sequence for x in the path-counting mode is 012120120120 , giving $m = 2$ and p -value 0.04. The only optimal state sequence for x in the deterministic mode is 012020120120 , giving $m = 4$ and p -value 0.30.

Theorem 22 immediately gives an interesting corollary.

Theorem 23. There is an x such that the explanation of x in deterministic mode and the explanation of x in path-counting mode are disjoint.

See Figure 4 for illustration of Theorems 22 and 23.

Theorem 24 (proved by computer). The optimal number of states for $x = 0110001000$ in the path-counting mode is 4, corresponding to $m = 2$ and a p -value of 0.79. The optimal number of states for $x = 0110001000$ in the word-counting mode is 2, corresponding to $m = 7$ and a p -value of 0.6.

The corollary we seek is now immediate from Theorem 24.

Theorem 25. There is an x such that the explanation of x in word-counting mode and the explanation of x in path-counting mode are disjoint.

See Figure 5 for an illustration of Theorems 24 and 25.

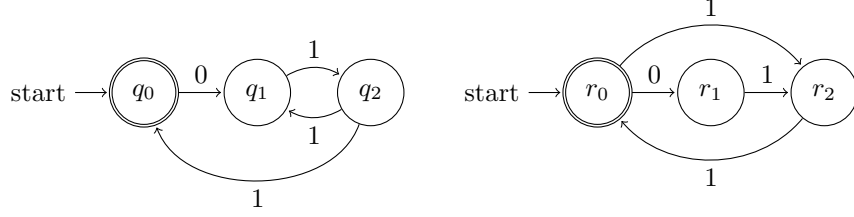


Figure 4: Optimal models for 01111011011 in the path-counting (left) and deterministic (right) modes (Theorem 23).

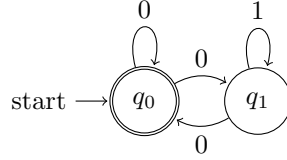


Figure 5: An optimal model for $x = 0110001000$ in the word-counting mode (Theorem 25). Note the use of multiple paths.

4 Determinism and automatic complexity

In [3] we give an example of a word x such that $A^-(x) - A_N^\pi(x) = 2$. We conjecture that the differences $A^-(x) - A_N^\pi(y)$ are unbounded as $|y| \rightarrow \infty$. However, for most words, the difference between A^- and A_N^π is small. Let $[b] = \{0, \dots, b-1\}$ and let $|X|$ denote the cardinality of a set X . We show that most words have A^- -complexity at most $(\frac{1}{2} + \varepsilon)n$ in the following sense.

Theorem 26. *For each $\varepsilon > 0$ and integer $b \geq 1$,*

$$\lim_{n \rightarrow \infty} b^{-n} \left| \left\{ x \in [b]^n : \frac{A^-(x)}{n} \leq \frac{1}{2} + \frac{1}{2b} + \varepsilon \right\} \right| = 1.$$

Proof sketch. The idea is *derandomization*, or perhaps more accurately *determinization*, of Kayleigh graphs (Figure 1). Whenever there is a state with nondeterministic out-behavior, split it into two states as in Figure 6. This will only happen about a fraction $\frac{1}{b}$ of the time, so the total number of states will be about

$$\frac{n}{2} + \frac{n}{2} \cdot \frac{1}{b} = \left(\frac{1}{2} + \frac{1}{2b} \right) n.$$

By the Law of Large Numbers, the statement of the Theorem follows. \square

We also know that A_N^π and A_N^ω have the same sharp upper bound. The argument in [2], to the effect that $n/2 + 1$ is sharp, applies to them equally.

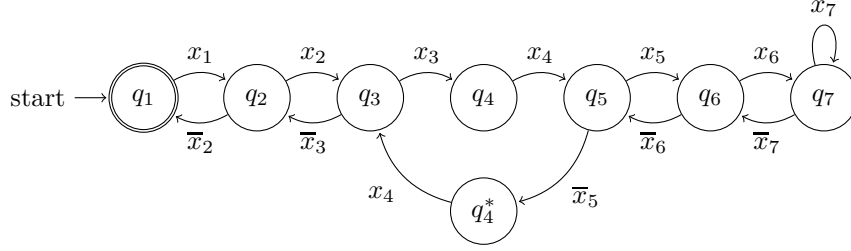


Figure 6: A deterministic finite automaton that only accepts one word $x = x_1x_2x_3x_4x_5x_6x_7\bar{x}_7\bar{x}_6\bar{x}_5x_4\bar{x}_3\bar{x}_2$ of length $n = 13$. It is obtained by “exploding” the state q_4 in a Kayleigh graph (Figure 1).

5 Automatic complexity of doubletons

Definition 27. The word-based automatic complexity $A_N^\omega(\mathcal{F})$ of a finite set $\mathcal{F} \subseteq \{0, 1\}^n$ to be the minimum number of states of an NFA M such that

$$L(M) \cap \{0, 1\}^n = \mathcal{F}.$$

The path-based automatic complexity $A_N^\pi(\mathcal{F})$ is the minimum number of states of an NFA M such that in addition M has only $|\mathcal{F}|$ many accepting paths of length n .

This generalizes automatic complexity from the case where \mathcal{F} is a singleton. Clearly $A_N^\omega \leq A_N^\pi$. We shall see in Theorem 30 that $A_N^\omega(\mathcal{F}) \neq A_N^\pi(\mathcal{F})$ in general when $|\mathcal{F}| = 2$. We conjectured in Conjecture 13 that $A_N^\omega(\mathcal{F}) \neq A_N^\pi(\mathcal{F})$ for some \mathcal{F} with $|\mathcal{F}| = 1$.

Theorem 28 (Hyde [2] ($f = 1$), Chambers [1] ($f = 2$)). *The automatic complexity of a set $\mathcal{F} \subseteq \{0, 1\}^n$ (with one accept state allowed) of size f satisfies*

$$A_N^\pi(\mathcal{F}) \leq f \lfloor n/2 \rfloor + 1.$$

Proof sketch. The proof is easy upon consideration of Figures 1 and 7. □

Theorem 29 (proved by computer). *Let $\mathcal{F} = \{0110, 1111\}$. For any NFA M such that $L(M) \cap \{0, 1\}^4 = \mathcal{F}$, M has at least 3 states, and if M has 3 states then M has at least 3 accepting paths of length 4.*

Theorem 30. *There is an n and a finite set $\mathcal{F} \subseteq \{0, 1\}^n$ such that*

$$A_N^\omega(\mathcal{F}) \neq A_N^\pi(\mathcal{F}).$$

Proof. Let $n = 4$ and let $\mathcal{F} = \{0110, 1111\}$. By Theorem 29, the automaton C in Figure 8 is optimal. Thus

$$A_N^\omega(\mathcal{F}) = 3 < 4 \leq A_N^\pi(\mathcal{F}). \quad \square$$

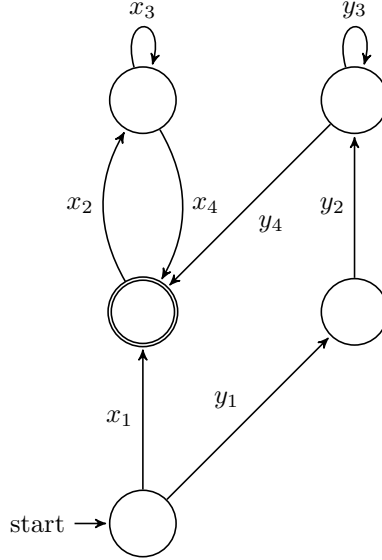


Figure 7: An automaton witnessing the Chambers–Hyde bound (Theorem 28) for $n = 4$ and $f = 2$.

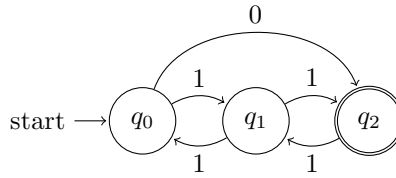


Figure 8: The optimal automaton C with $L(C) \cap \{0, 1\}^4 = \{0110, 1111\}$ by necessity accepts 1111 along two paths, giving Theorem 30.

Acknowledgments

We thank Greg Igusa for contributing ideas to Section 2. This work was partially supported by a grant from the Simons Foundation (#315188 to Bjørn Kjos-Hanssen). This material is based upon work supported by the National Science Foundation under Grant No. 1545707.

References

- [1] Patrick Chambers. Automatic complexity of doubletons. Verbal presentation as a final exam for MATH 657, University of Hawaii at Mānoa, Spring 2014.
- [2] Kayleigh Hyde and Bjørn Kjos-Hanssen. Nondeterministic automatic complexity of overlap-free and almost square-free words. *Electron. J. Combin.*, 22(3):Paper 3.22, 18 pp., 2015.
- [3] B. Kjos-Hanssen. Shift registers fool finite automata. *ArXiv e-prints*, July 2016.
- [4] Bjørn Kjos-Hanssen. Kolmogorov structure functions for automatic complexity. *Theoret. Comput. Sci.*, 607(part 3):435–445, 2015.
- [5] Jorma Rissanen. *Information and complexity in statistical modeling*. Information Science and Statistics. Springer, New York, 2007.
- [6] Jeffrey Shallit and Ming-Wei Wang. Automatic complexity of strings. *J. Autom. Lang. Comb.*, 6(4):537–554, 2001. 2nd Workshop on Descriptive Complexity of Automata, Grammars and Related Structures (London, ON, 2000).
- [7] Ludwig Staiger. The Kolmogorov complexity of infinite words. *Theoret. Comput. Sci.*, 383(2-3):187–199, 2007.
- [8] Nikolai K. Vereshchagin and Paul M. B. Vitányi. Kolmogorov’s structure functions and model selection. *IEEE Trans. Inform. Theory*, 50(12):3265–3290, 2004.